Classification of Power Quality Events using Wavelet Packet Transform and Extreme Learning Machine

Chirag Naik*, Faizal Hafiz†, Akshya Swain‡ and A. K. Kar†

*Sarvajanik College of Engineering & Technology, Gujarat, India.
Email: naik.chirag@gmail.com
†Department of Electrical & Computer Engineering, The University of Auckland, Auckland, New Zealand
Email: faizalhafiz@ieee.org, a.swain@auckland.ac.nz
‡Department of Electrical Engineering, Unitec Institute of Technology, Auckland, New Zealand
Email: akar@unitec.ac.nz

Abstract—A novel method of classifying Power quality (PQ) events using Wavelet Packet Transform (WPT) and Extreme Learning Machines (ELM) has been proposed. In recent times, the power quality has been a major research concern due to changing regulations, liberalized distribution market and increased use of power electronic based equipment. The first step of any remedial action requires proper identification of PQ events. One of the major challenges of this event identification is to extract significant features from the limited measurements, which can subsequently be used for the classification. Therefore, in the present study Wavelet Packet Transform (WPT) has been used to obtain several mathematical features. These features can segregate both single and simultaneous PQ event occurrences. Further to improve the classification performance, the ELM based classifier has been used. This classifier significantly reduces the training time by many-fold. The performance of the proposed approach has been compared with ANN based classifier considering over 1000 PQ signals from various PQ events. The results of the simulation demonstrate that the proposed approach can achieve over 99% classification accuracy.

Keywords—Classification; power quality; signal processing; power distribution; extreme learning machines

I. INTRODUCTION

The power quality (PQ) has become a major concern for the utilities in the past few years with the changing landscape of the electrical power utilization. Especially, the increasing use of electronic equipment and consequent many-fold rise in non-linear loads have made the maintenance of PQ a challenging task. The partial or complete failure of sensitive equipment, data processing errors and data losses are some of the detrimental effects caused by poor PQ. Further, the revised regulations and liberalization of distribution market have highlighted the importance of maintaining good PQ [1].

The consumer load is the major factor which contributes to the PQ deterioration in the changed load scenario. At the same time, the utility operations such as load switching, fault clearing, maintenance activities also contribute to PQ deterioration. Hence, the responsibility of maintaining good PQ must be shared by, both, utility and its consumers.

The remedial actions for PQ improvement can only be initiated, if the nature the PQ disturbances, their impact on overall PQ and the frequency of their occurrence is correctly monitored [2]. Hence, the identification of PQ disturbance through classification is essential before taking any remedial action [3]–[5]. For this purpose, the IEEE 1159-2009 standard broadly defines the different PQ disturbances [6] such as voltage sag, voltage swell, interruptions, harmonics, oscillatory and impulsive transients on the basis of their magnitude, duration and spectral content.

The major challenge of the PQ classification is to identify wide variety of PQ disturbances from limited measurements. However, it is not possible to classify PQ disturbance/event directly from the acquired signals owing to wide variations in frequency (from fundamental to several kHz), duration (few cycles to several seconds) and magnitude. To overcome this problem, usually several mathematical features are derived from the acquired PQ signals through suitable signal processing technique [7]–[9]. Over the years, several signal processing techniques such as Fourier Transform (FT), Short time Fourier transform (STFT), Wavelet Transform (WT) and Wavelet packet transform (WPT) have been utilized for the feature extraction.

The Fourier Transform (FT) is suitable for the analysis of stationary harmonics, and has been used quite effectively for harmonic analysis. However, FT completely transforms the amplitude-time representation to magnitude-frequency representation. In both representations; either time or frequency information is lost. For this reason, FT is not suitable for the analysis of non-stationary PQ disturbances [10]. The windowed version of FT, Short time Fourier transform (STFT), can retrieve both time and frequency information. However, due to fixed window width, the time resolution is compromised [11]. The Wavelet Transform (WT) is capable of fetching both time and frequency information. For this reason, most of the recent approaches [8], [12]–[14] rely on WT. In this work, a variant of WT, Wavelet packet transform (WPT), has been used to extract the mathematical features from the PQ signals.

To classify the PQ events using extracted features any intelligent classifier can be used, considering the extracted features as pre-cursors and PQ event type as outputs. e.g., Artificial Neural Networks (ANN) [15], [16], Fuzzy Logic Controllers [17], Support Vector Machines (SVM) [18], Rule-Based systems [19], Hidden Markov models [20] etc.

Note that the overall classification accuracy is dependent on both, extracted features and classifier. Hence, to effectively classify/identify the PQ events it is essential to combine effective distinguishing features with a robust classifier. For
this purpose in this work, ten mathematical features extracted using WPT have been proposed. The proposed features can be used as pre-cursors to any intelligent classifier for the final classification. For this purpose, the present study proposes the use of Extreme Learning Machine (ELM) [21]–[24]. The extreme learning machine is in essence a Single-hidden Layer Feed-forward Network (SLFN), with simplified yet quite effective learning algorithm. The use of ELM as a classifier enables significant reduction in training time while maintaining higher classification accuracy. To evaluate the effectiveness of the proposed, WPT-ELM approach, test data sets containing more than 1000 signals were derived following the definitions of IEEE 1159-2009. Further, the performance of ELM is compared with ANN using the proposed mathematical features on single and multiple simultaneous disturbances. The rest of the article is organized as follows: proposed feature extraction technique using WPT is explained in Section II. The ELM based classifier is briefly explained in Section III. The results are illustrated in Section IV with conclusions in Section V.

II. THE WAVELET PACKET TRANSFORM

The acquired PQ signals are time domain signals which give amplitude-time representation. WT can simultaneously fetch both, time and frequency, information necessary for the feature extraction. However, the complex computation involved with the WT often makes it unfeasible for the practical applications. The discrete wavelet transform (DWT) reduces the computational complexity of WT and hence, utilized in many PQ analysis. Note that the signal processing technique used in this work i.e. WPT, is essentially an extension of DWT.

In both, DWT and WPT, the signals are decomposed at several levels. At each level, signals are decomposed by half frequency high pass and low pass filter. The output signals from the high-pass filter are referred to as details, \( cD_j \) and the same from the low-pass filter as approximations, \( cA_j \), respectively. At every level of decomposition the signal is down-sampled by a factor of 2. In DWT at each level approximations, \( cA_j \) are discarded and only details, \( cD_j \) are passed to the next level for further decomposition. On the other hand, in WPT both details and approximations are preserved and analyzed at each level of decomposition. Thus, WPT can simultaneously analyze both high and low frequency content of the time series. This is desirable as voltage sag, swell and interruptions are related with low frequencies while oscillatory transients and harmonics contain medium and/or higher frequencies.

Fig. 1 shows the WPT algorithm with three levels of decomposition. The original PQ signal is at level (0,0). The details and approximations of the original PQ signal at the first level of decomposition are denoted as (1,0) and (1,1). At level (0,0) the sampled PQ signal is passed through first level of half band high pass and half band low pass filters to get detail cD(1,0) and approximation cA(1,1) and the process is continued further till the third level of decomposition.

III. PQ SIGNAL ACQUISITION AND FEATURE EXTRACTION

The PQ disturbance signals were acquired by MATLAB simulations following the broad definitions of IEEE 1159-2009 standards. To test the proposed technique, several multiple simultaneous disturbances were also considered, e.g., sag with harmonics, swell with harmonics, sag with transients and swell with transients. All the simulated signals were normalized with respect to 100 V peak to peak pure sinusoidal signal. The simulated disturbances are illustrated in Fig. 2. For each disturbance class shown in Table I, several cases were simulated following IEEE 1159. For example, several cases of voltage sag varying from (10%-90%) in magnitude and 1 cycle to 15 cycle in duration were included. Similarly, simulated momentary interruption varied from 1 cycle to 15 cycles. The variation in oscillatory transients was induced by using different values of capacitance from 100 \( \mu \)F to 5000 \( \mu \)F, switching instances and durations. The harmonic signals were simulated using various sources of harmonics, which include: inverters with sinusoidal pulse width modulation scheme, uncontrolled rectifier and controlled rectifiers with various firing angles. These simulated signals were used to create a test data set containing 1152 signals which was utilized for investigating the performance of the proposed approach.

The feature extraction plays an important role in the classification. The extracted features should mathematically convey the information about the respective PQ disturbance signal. In this paper, for each PQ disturbance signal, 10 discriminating mathematical features were computed based on the WPT decomposition at different levels as shown in Table II. As seen in Fig. 3, the computed WPT based features are capable of distinguishing between the classes, although their values in some cases overlapping. This further underlines the usage of an intelligent classifier for the accurate classification.

![WPT algorithm with three level of decomposition](image-url)
of the PQ events. In the next section, the proposed classifier based on ELM is explained.

IV. EXTREME LEARNING MACHINE BASED CLASSIFIER

An Extreme Learning Machine (ELM), proposed by Huang et. al. [21], is in essence a Single-hidden-layer Feed-forward Network (SLFN) albeit with a simplified learning mechanism. In conventional Neural Network (NN) learning approaches, the target function is approximated by adapting network parameters (connection weights and biases of neurons) as per learning data. However, this rigid approach to adjust each network parameter is the main cause of NN’s inability to use non-differentiable activation functions and pro-longed training durations [21]–[24].

To overcome this problem, a simplified learning approach is used for ELM. Huang et. al [21] demonstrated that it is possible to set the input weights and biases of the hidden layer neurons in SLFN without compromising its generalization capabilities. With input weights and biases to hidden layer determined randomly, only the connection weights from hidden to output layer need to be evaluated; which can be done analytically. In other words, with this learning approach only small fraction of the network parameters have to be evaluated for given learning data which leads to significant reduction in a training time and a better generalization capability for unseen test data.

The simplified learning approach for ELM can be explained by a simple problem consisting of n inputs, m outputs and N training samples. For this problem, ELM will have n and m neurons in the input and output layers. Assuming the size of
the hidden layer (hidden neurons) to be \( \hat{N} \), the output of ELM for the \( j^{th} \) training sample \( (o_j) \) can be given by the following equation [21], [22],

\[
\sum_{i=1}^{\hat{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\hat{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, \tag{1}
\]

with \( j = 1, \ldots, N, w_i = [w_{i1}, \ldots, w_{im}]^T \) and \( \beta_i = [\beta_{i1}, \ldots, \beta_{im}]^T \)

where, \( g_i(x) \), \( w_i \) and \( b_i \) are the activation function, weight vector and bias of the \( i^{th} \) hidden neuron, respectively. On the basis of (1), the output of the hidden and the output layer can be represented by (2) and (3), respectively.

\[
H = \begin{bmatrix}
g(w_1 \cdot x_1 + b_1) & \ldots & g(w_{\hat{N}} \cdot x_1 + b_{\hat{N}}) \\
\vdots & \ddots & \vdots \\
g(w_1 \cdot x_N + b_1) & \ldots & g(w_{\hat{N}} \cdot x_N + b_{\hat{N}})
\end{bmatrix}_{N \times \hat{N}}
\]

\[
\beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_{\hat{N}}^T
\end{bmatrix}_{\hat{N} \times m} \quad \text{and} \quad T = \begin{bmatrix}
t_1^T \\
\vdots \\
t_N^T
\end{bmatrix}_{N \times m}
\]

\[
H \beta = O \tag{3}
\]

Note that, the input weights and bias of the hidden neuron \( (w, b) \) are randomly determined. Hence, output of the hidden layer \( (H) \), can be evaluated easily using (2). The next step is to determine the connection weights from hidden to the output layer \( (\beta) \) by solving following analytical equation [21]–[24],

\[
H \beta = T \tag{4}
\]

In this work, an updated version of ELM algorithm [24], is used to evaluate \( \beta \) which is given by (5). The size of hidden layer \( \hat{N} \), and the constant, \( C \), in (5) were determined heuristically to be 200 and 215.

\[
\beta = H^T ( \frac{I}{C} + HH^T )^{-1} T \tag{5}
\]

V. Results and Discussions

The performance of the proposed WPT-ELM based classifier is evaluated using the test data containing 1152 signals. For each of the signal, ten distinguishing features were evaluated using WPT and mathematical function listed in Table II. Out of 1152 data samples, 864 samples were (75% of data) randomly selected for the training of the classifier and the remaining 288 samples were used for validation. In addition to ELM, ANN (feed-forward network with 10 hidden neurons) based classifier

<table>
<thead>
<tr>
<th>Features</th>
<th>Function</th>
<th>WPT Decomposition Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>Energy</td>
<td>( \sum_{n=1}^{N} (x_1^n + x_2^n + \cdots + x_m^n) )</td>
</tr>
<tr>
<td>( f_2, f_3 )</td>
<td>Minimum</td>
<td>( \min_{1 \leq n \leq N} x_n )</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>Maximum</td>
<td>( \max_{1 \leq n \leq N} x_n )</td>
</tr>
<tr>
<td>( f_5 )</td>
<td>Variance</td>
<td>( \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})^2 )</td>
</tr>
<tr>
<td>( f_6 )</td>
<td>Standard Deviation</td>
<td>( \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})^2} )</td>
</tr>
<tr>
<td>( f_7, f_8 )</td>
<td>Skew</td>
<td>( \frac{1}{N} \sum_{n=1}^{N} \left( \frac{x_n - \bar{x}}{\sqrt{\sum_{n=1}^{N} (x_n - \bar{x})^2}} \right)^3 )</td>
</tr>
<tr>
<td>( f_9, f_{10} )</td>
<td>Kurtosis</td>
<td>( \frac{1}{N} \sum_{n=1}^{N} \left( \frac{x_n - \bar{x}}{\sqrt{\sum_{n=1}^{N} (x_n - \bar{x})^2}} \right)^4 )</td>
</tr>
</tbody>
</table>

Fig. 3: Co-relation of Extracted Features with the PQ event Classes (a) \( f_1 \) and \( f_{10} \), (b) \( f_4 \) and \( f_5 \), (c) \( f_7 \) and \( f_8 \)
TABLE III: PERFORMANCE COMPARISON OF ELM AND ANN BASED CLASSIFIERS - I

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best</td>
<td>Mean</td>
</tr>
<tr>
<td>ELM</td>
<td>99.31%</td>
<td>98.06%</td>
</tr>
<tr>
<td>ANN</td>
<td>98.62%</td>
<td>97.22%</td>
</tr>
</tbody>
</table>

was also used for the performance comparison. Since, the input parameters of ELM’s hidden layer and initial parameter of ANN are selected randomly, both the classifiers were trained 25 times using the same training data. For both ELM and ANN, the architecture with the best accuracy (out of 25 runs) was further used to evaluate the performance using validation data. The results of the classifier training are listed in Table III. As expected, ELM outperforms ANN; the training time for ELM is almost 19 times less than that for the ANN. In addition, the standard deviation is comparatively less in ELM which indicates stable performance.

The prediction of classifiers for the validation data is shown in Table IV. Both the classifiers identified test instances of Class 4 - 10 with 100% accuracy. Note that PQ instances in Class 6-10 are simultaneous disturbances; 100% accuracy in identification of these instances indicates effectiveness of the extracted features. Mis-classification was observed for the PQ instances of Class 2 and 3 (Sag and Swell) albeit its very small and within acceptable limits. Note that ANN based classifier failed to identify the Class 1 events which further underlines the role of classifier in the overall performance. Overall, with WPT-ELM based approach gives classification accuracy of 99.3%.

VI. CONCLUSIONS

The Power Quality classification using on WPT-ELM combination is proposed. The efficacy of the proposed approach was evaluated on ten distinct, single and multiple simultaneous disturbances which were simulated following the definitions of IEEE 1159-2009 standard. Ten distinct mathematical features were extracted using WPT which can be used as pre-cursors to any classifier. To enhance the classification accuracy further, an alternate classifier based on ELM is proposed and its performance is compared with ANN based classifier. The results indicate that the simplified learning algorithm of ELM leads to significant reduction in training time while providing enhanced classification accuracy on unseen data owing to better generalization capabilities.

REFERENCES


TABLE IV: PERFORMANCE COMPARISON OF ELM AND ANN BASED CLASSIFIERS - II

<table>
<thead>
<tr>
<th></th>
<th>ELM Predictions</th>
<th>ANN Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>96.9%</td>
<td>3.1%</td>
</tr>
<tr>
<td>3</td>
<td>97.4%</td>
<td>2.6%</td>
</tr>
<tr>
<td>4</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>100%</td>
<td>-</td>
</tr>
</tbody>
</table>

Overall 99.3% 0.7% 98.6% 1.4%


